地域研究のための資源共有化システムとメタデータに関する研究

AUTHOR(S):

原, 正一郎

CITATION:

原, 正一郎. 地域研究のための資源共有化システムとメタデータに関する研究. 東南アジア研究 2009, 46(4): 608-645

ISSUE DATE:

2009-03-31

URL:

http://hdl.handle.net/2433/88029

東南アジア研究　46巻4号　2009年3月

地域研究のための資源共有化システムとメタデータに関する研究

原　　正一郎*

Studies on Resource Sharing System and Metadata for Area Studies

HARA Shoichiro*

Area informatics is the new paradigm in area studies to facilitate accumulation and creation of knowledge in areas. In the process, a huge variety of databases such as catalogs, archives, full texts, images, movies, sounds, statistics, and so on, are being organized and published on the Web; these will be the sources of area-specific knowledge. However, it is difficult for researchers to find and access appropriate databases to retrieve resources effectively because each database is independent and dispersed on the Web; furthermore, their data structures and retrieval procedures are different.

Resource Sharing System, an outcome of area informatics, is an innovative information retrieval system that has been developed to solve such problems. It is a server-side system that hides from users each database system's particular data structures and retrieval procedures by employing standard metadata and standard retrieval protocols.

In this paper, area informatics is introduced through a brief overview of the relationship between area studies and information sciences. After discussing the structure of Resource Sharing System, the new notion of "metadata suites" is introduced and explained. This is a guideline to build databases to be included in Resource Sharing System. Finally, a sample metadata compiled by CIAS is presented and its availabilities discussed.

Keywords: area informatics, resource sharing system, metadata, metadata-suites
キーワード: 地域情報学，資源共有化システム，メタデータ，メタデータ・スイート

I　は　じ　め　に

本稿では，地域研究の新しい研究パラダイムである地域情報学（area informatics）において重要な役割を担いつつある資源共有化システム（resource sharing system），および資源共有

* 地域研究統合情報センター; Center for Integrated Area Studies, Kyoto University
e-mail: shara@cias.kyoto-u.ac.jp

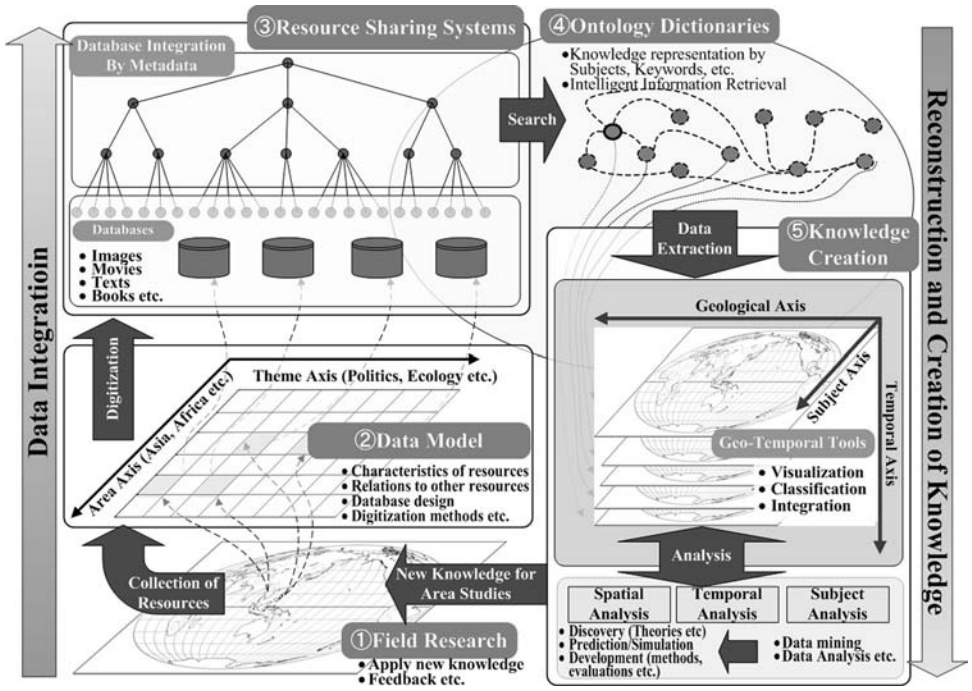

図 1　データ・情報・知識のフローから見た地域情報学モデル

化システムの基盤となるメタデータ（meta data）[注1)] について，京都大学地域研究統合情報センター（以下では地域研）における研究・開発の現状を報告するとともに，今後の展開を目指した提案を行う。

地域情報学は，地域研究に関するデータや情報の共有，地域の解明と理解，さらに地域研究知の創生を支援する情報学的パラダイムである［柴山・原 2008: 28–35］。地域研究と情報学の関連を俯瞰するために，地域研究のプロセスをデータ・情報・知識のフローという観点から整理した，地域情報学モデルを図 1 に示す［原・柴山 2007: 71–78］。

地域研究はフィールドから始まる（図 1–①）。フィールドではインタビュー・観察・資料調査などの研究活動が展開される。その成果として，多種多量の静止画・動画・音声・メモ・標本・文献・地図・数値データなどの資料が収集・生成される。

これらの資料は，データモデルの検討を通じて個別のデータベースとして組織化される（図 1–②）。データモデルの検討では，資料のメディア特性に応じて適切なデジタル化の手法を決定する。同時に，資料管理と検索に適したメタデータの設計，つまりデータ項目・名称・粒度・言語・符号化法などの決定と記述規則の策定を行う。データモデルの検討結果に従って資料のデジタル化とメタデータの作成を行う。デジタルデータとメタデータからデータベースを構築する。あらゆる情報処理にとってデータベースはまさに基地であり，その設計と実装の良

し悪しが，以降の情報処理の速度・精度・利用法などに影響を与える。

個別のデータベースは資源共有化システムに統合される（図1-③）。資源共有化システムとは，インターネット上の多種多量なデータベースを同時に検索し，それらの検索結果を適切にまとめて利用者に提示する情報システムである。例えば，地域研の資源共有化システムに対して「canal」という文字列で検索を要求すると，英国議会資料地図データベースなど地域研が公開しているデータベースに加え，国立民族学博物館や総合地球環境学研究所のデータベースなども同時に検索し，canal に関する様々な資料の所在情報や画像などを表示する。[1)]

データの記述言語や語彙はデータベースごとに異なっており，資源共有化システムを有効活用するには，言語や語彙の違いを乗り越えた検索・知識体系の枠組みが必要である。その枠組みとしてオントロジーに注目している（図1-④）［溝口 2005］。例えば，地名には変遷・地理的包含関係・同名異地などの問題があるため，地名に関するシソーラスと地理参照を備えたデジタル地名辞書が必要となる。海外では Alexandria Digital Library の *Gazetteer Development* ［Alexandria 2004］や Getty Thesaurus of Geographic Names Online［Getty 2009］など大規模なデジタル地名辞書がいくつか公開されているが，[2)] 日本には同様の辞書がないので，資源共有化システムの研究の一環としてデジタル地名辞書の構築に着手した［桶谷 2007: 79–86］。

知識生成では資源共有化システムからデータを検索し，それらを分析・統合・解析し，新しい仮説形成や知識発見を試みる（図1-⑤）。具体的には可視化・語彙分析・統計・時空間などの基本的なデータ分析，データ間の関連付け，分類・因果関係の抽出・予測などを行う。その成果は再びフィールドへフィードバックされ，新しい研究プロセスが始まる。このような知識生成を支援するため，HuMap と呼ばれる空間情報処理ツール［原 2008: 128–135］や HuTime と呼ばれる時間情報処理ツール［関野・久保 2007: 183–188; 関野 2008: 140–148］などの開発も進めている。

本稿では，地域研究におけるデータ・情報・知識の循環を支援する情報学的体系の確立と，そこで必要となる手法の開発を，地域情報学の目標と捉えている。とりわけ資源共有化システムは，地域研究に関するデータ・情報・知識を統合するための要となる情報システムである。Ⅱ章において，地域研究における資源共有化システムの枠組み，開発の経緯およびシステムの構造について述べる。

ところで，多様な情報を一括検索する手法としては，Google や Yahoo などの検索エンジンが普及している。これらの検索エンジンは，構造化[注2)]がそれほど高度ではない文書情報を文

1）本稿の執筆時点において，地域研の資源共有化システムは外部データベースと接続していない。しかし資源共有化システムのメカニズムは人間文化研究機構のシステムや国立国会図書館の PORTA と同じであり，接続実験を予定している。

2）ここで例示した海外のデジタル地名辞書は現在の地名に関する辞書である。東南アジアの地名についても，ある程度は収容されている。

字列により検索する手段である。構造化の水準が低いため，例えば「近衛」について検索しても，それが人名なのか，地名なのか，あるいは職名であるのかといったデータの意味を区別できず，その結果として大量の検索ノイズが混入する。これに対してデータベースでは，レコード中の各フィールドが主題・人名・地名など意味を表しているので，データの意味を指定した検索が可能である。[注3)] 研究者にとって，検索結果に関する場所・時間・人物・事象・事件などの区別は重要である。したがって，資源共有化システムは高度に構造化されたデータベースでなければならない。Ⅲ章は本稿の中核であり，資源共有化システムで利用するメタデータの基本的な構造・特徴・利用法・問題点について整理した後，資源共有化システムの新しいデータモデルであるメタデータ・スイート（Metadata Suites）と，それに関連するメタデータについて詳述する。

最後にⅣ章で資源共有化システムの今後について検討する。

II　資源共有化システムの概要

資源共有化システムは，地域研究に関する様々なデータ・情報・知識を統合するための要となるシステムである。これまでのデータ共有化では，書誌などデータ構造が似ているデータベース同士の統一を目指していた。資源共有化システムは，これらとは一線を画した斬新な情報システムであり，図書館・アーカイブ・博物館など，資料の整理体系やデータ構造が異なるデータベースの連携を目指している。本章では資源共有化システムの枠組み，開発経緯，およびシステムの構造について述べる。

II-1　資源共有化システムの枠組み

地域研のデータベースは，前述のデータモデルの検討結果に基づいて構築されている。ところが，殆どのデータベースは地域や研究テーマごとに異なるデータモデルを採用していたため，データ構造や検索手順は共通化されていない。学際的研究を展開する第一歩は，関連する資料を多くのデータベースから収集することであるが，このままではデータベースを１つずつ検索しなければならず，極めて非効率的である。[注4)] 資源共有化システムでは，異なるデータベースをシームレス（seamless）に接合することを目指している。シームレスとは，インターネット上に分散している複数のデータベースを，仮想的な単一データベースに見せる仕掛けである。本研究では，このような情報システムを総称して，資源共有化システムと呼ぶことにする。

データを共有化する初期の方法は，全てのデータベースのメタデータと検索手順を同じものに作り替える，いわゆる「データベースの統一」であった。これは全国書誌ユーティリティや

初期の病院情報システムなどで試みられた方法である。

全国書誌ユーティリティについては，基本的に書誌という同質のメタデータが対象であったため，国立情報学研究所［2009］や OCLC［2009］などのような成功例が見られた。しかし図書館とアーカイブの共有化を考えた場合，たとえ両者が同じ文書資料を扱っていたとしても，それぞれの資料整理の体系が異なっているため，これらのメタデータを統一することは困難である。病院情報システムの場合も，語彙や単位やカルテの書法などが診療科あるいは病院ごとに異なっているため，メタデータについての合意を形成することは困難であった。そのため，全国規模の病院情報システムは成功しなかった。つまり専門性の異なるデータベースのメタデータを統一することは事実上不可能であった。

大学の研究所などが公開しているデータベースを眺めてみると，同じ研究分野のデータベースであるにも拘わらず，メタデータはバラバラであることが多い。データベース作成者が標準メタデータの存在を知らないあるいは不勉強である，研究上の必要性から特殊なメタデータを定義した，資料の内容的・物理的特性のために通常とは異なる標準メタデータを採用した，などの理由が考えられる。

さらに，既存のデータベースに手を加えようとすると，担当部署の反発を受けることが多いことも経験的な事実である。これはデータベース運用歴の長い組織に顕著であるが，既に業務としての手順が確立しているので変更したくない，あえて標準メタデータに変更する必要性を感じていない，変更作業に膨大な時間と資金がかかる，などが理由となっている。以上のような状況から，データベースの統一は共有化の主要な方法ではなくなっている。

前述の検索エンジンは，一度作成すれば変更を加えることの少ない静的文書，例えば HTML（Hyper Text Markup Language）で書かれたホームページの検索に適していると言われている。これに対してホームページからリンクされているデータベースは，データ構造や検索手順がそれぞれ異なっているため，通常の検索エンジンで利用することが困難である。[注5)]

データベースの統一や検索エンジンとは異なる共有化法として「交換用メタデータ」がある。これはデータ構造の異なる情報システム間でデータ交換を行うための主要な方法となっている。最初に，各情報システムの固有メタデータから独立したデータ交換用メタデータを定義する。次に，システム固有のメタデータを交換用メタデータに変換するプログラムと，交換用メタデータをシステム固有のメタデータへ変換するプログラムを，各情報システムにおいて一組ずつ用意する。データを送信する場合，各情報システムはシステム固有のメタデータを交換用メタデータに変換してから送信する。データを受信する場合，情報システムは受信した交換用メタデータをシステム固有のメタデータに変換する。各情報システムでは固有メタデータを変更する必要がなく，またデータの送受信先がどこであっても一組の変換プログラムで対応できるという特徴がある。この例として，企業間取引のための電子化ビジネス文書の交換規約

EDIFACT［ISO 1987］や医療情報交換のための標準規約 HL7［2007］などがある。

資源共有化システムでも同様の方法を採用している。資源共有化システムの場合，各データベースの固有メタデータから独立した共有化用のメタデータを定義する。次に，各データベースの固有メタデータを共有化用のメタデータに変換し，新たにメタデータベースを構築する。検索はこのメタデータベースに対して行う。資源共有化システムの構造についてはⅡ-3節で述べる。

II-2　資源共有化システム開発の経緯

資源共有化システムの開発は2000年から国文学研究資料館を中心に開始された［原・安永 2000: 25-52; 原 2002a: 17-35］。このシステムは国文学研究資料館内のデータベースを共有化することが主眼であり，機関外データベースとの共有化は大阪市立大学学術情報総合センターなどに限定されていた。

本格的な異分野データベース間の共有化実験は2002年から開始された。[3] このときの研究目的は，参加機関の25個のデータベースを連携させるために必要な検索手順とメタデータを確立することであった［原 2002b: 968-974; 原 他 2003: 17-22］。国文学研究資料館の資源共有化システムを基盤として共有化は実現できたものの，システムの動作が不安定であり，[4] 一般公開には至らなかった。

国文学研究資料館を含む５つの人文系大学共同利用機関を母体として，2004年に大学共同利用機関法人人間文化研究機構が発足し，これに合わせて「人間文化研究総合推進事業――人間文化研究資源共有化推進事業」が開始された。この事業には「研究資源共有化システム」と呼ばれる資源共有化システムの研究・開発も含まれており，国文学研究資料館および総合研究大学院大学における研究成果が継承されることになった。以下では人間文化研究機構による資源共有化システムを NIHU システムと呼ぶ。

NIHU システムの特色は，100個を超えるデータベースを安定的に相互接続すること，および時空間データ解析システムの開発である。NIHU システムは2008年４月から公開されている［原 2007: 107-136; 人間文化研究機構 2009］。NIHU システム以外の資源共有化システムとして，例えば国立国会図書館［2009］では PORTA の開発を積極的に進めている。PORTA のような機構外システムとの相互接続が NIHU システムの課題となっている。

地域情報学における資源共有化システムは，このような研究・開発の流れを受け継いでい

3）総合研究大学院大学を中心に国文学研究資料館，国際日本文化研究センター，国立民族学博物館，大阪市立大学，東京大学史料編纂所が参加した。

4）サーバ間の管理・調整機能がなかったため，例えば，ある機関の検索サーバが停止していると，そこで検索処理が停止してしまうなどの問題が発生した。

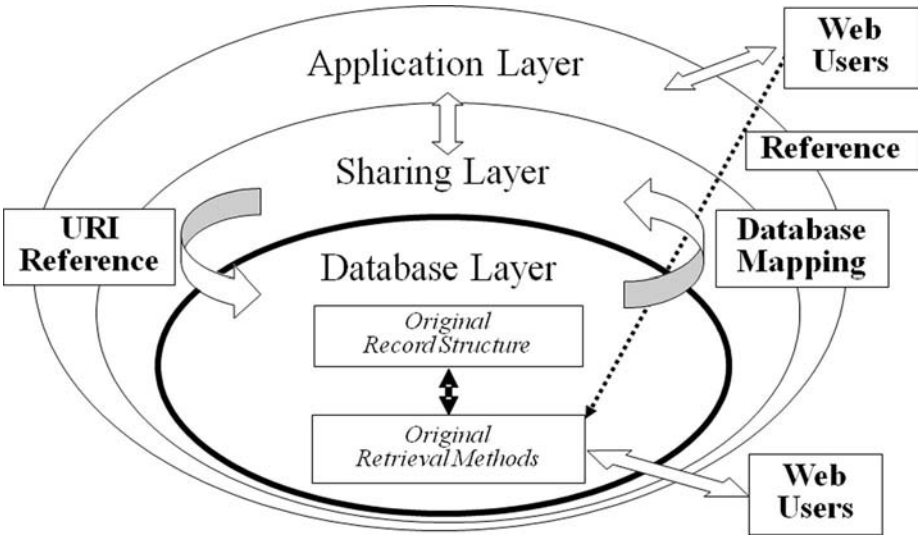

図 2　資源共有化システムのモデル

る。その実装例の一つが地域研の「地域研究資源共有化データベース（試行版）」である［地域研 2009］。以下では地域研究資源共有化データベースを CIAS システムと呼ぶ。CIAS システムの特色は，地域研究資料の特性を考慮し，より記述能力の高いメタデータを複数組み合わせていること，および時空間情報処理やオントロジーなどの最新の技術を積極的に取り込もうとしている点にある。

II-3　資源共有化システムの構造

資源共有化システムの基本的なモデルを図 2 に示す。資源共有化システムは 3 層の入れ子構造となっている。

資源共有化システムの一番内側の層を Database Layer と呼ぶ。ここには多様な個別データベースが該当する。前述のように，全てのデータベースのメタデータを統一することは困難である。したがって，資源共有化システムでは Database Layer に該当するデータベースのメタデータには手を付けない。

Database Layer の外側を Sharing Layer と呼ぶ。Database Layer に該当する各データベースのメタデータ構造がそれぞれ異なっていたとしても，意味的に共通したデータ項目が多くあることも経験的な事実である。そこで，多くのメタデータに共通な意味を持つデータ項目からなる新しいメタデータを定義する。これを本研究では共有化メタデータと呼ぶ。その上で，Database Layer に該当する各データベースのデータ項目を意味的に対応する共有化メタデータのデータ項目にマッピング（mapping）し，新たに共有化メタデータベースを構築する。こ

れが Sharing Layer の実体である。

資源共有化システムにおけるマッピングは，①元データベースの１データ項目を共有化メタデータの１つ以上のデータ項目へコピーする，②元データベースにおいて意味的に類似な複数データ項目を合併して共有化メタデータの１つ以上のデータ項目へコピーする，③共有化メタデータに対応しない元データベース中のデータ項目は無視する，の３種類である。多くの場合，マッピングにより元データベースの情報は劣化する。これから明らかなように，資源共有化システムにおけるデータ共有および検索の能力は，共有化メタデータの設計に大きく依存する。NIHU システムと CIAS システムのメタデータについてはⅢ章で詳述する。

なおマッピングに際し，共有化メタデータの各レコードには，マッピング元レコードへの URI（Uniform Resource Identifier）を参照情報として付与した。資源共有化システムでは Sharing Layer に該当する共有化メタデータベースを検索する。つまり利用者が資源共有化システムを Sharing Layer の外側から見ると，Database Layer に該当する全てのデータベースは共有化メタデータという単一のデータ構造となっている。

Sharing Layer の外側を Application Layer と呼ぶ。Sharing Layer に該当する各共有化メタデータベースのデータ構造は統一されているが，システムの実装法はバラバラである。例えば，ある共有化メタデータベースではテーブルと SQL を利用しているが，[注3)] 別の共有化メタデータベースでは XML（Extensible Markup Language）［W3C 2008］や SGML（Standard Generalized Markup Language）［ISO 1986］でマークアップ[注6)]された全文データを文字列探索エンジンにより検索しているかもしれない。つまり，色々な検索手順が可能であるため，このままではシームレスな検索は実現できない。そこで，データベースのハードウェアやソフトウェアに依存しない標準検索規約を実装して検索手順を統一した。これが Application Layer の実体である。資源共有化システムでは Application Layer が検索命令を処理する。つまり利用者が資源共有化システムを Application Layer の外側から見ると，Sharing Layer に該当する全ての共有化メタデータベースは単一の検索手順となっている。

情報検索を目的とした標準規約として Z39.50［ANSI/NISO 1995］を挙げることができる。[注7)] Z39.50 は欧米の図書館 OPAC（Online Public Access Catalog）間の相互検索で多く利用されている。しかし Z39.50 はインターネット環境に必ずしも適合した規約ではなく，また導入コストが高いなどの問題もあり，日本では余り普及していない。Z39.50 に替わり，最近では SRU/SRW（Search and Retrieve via URL/Search and Retrieve Web Service）［LC 2004a］が注目されている。[注8)] これはインターネット環境に適合した標準情報検索規約である。実装も Z39.50 に比べると容易かつ低コストである。今後は SRU/SRW が主流になると考えられるが，NIHU システムと CIAS システムでは，既に Z39.50 を採用しているデータベースとの連携も考慮し，Z39.50 と SRU/SRW の両方を実装している。

まとめると，資源共有化システムは，

- Sharing Layer の共有化メタデータにより，各データベースのデータ構造の相違を吸収する
- Application Layer の標準情報検索手順により，各共有化メタデータベースの検索手順の相違を吸収する

ことにより，インターネット上に分散しているメタデータと検索手順の異なる多種多量なデータベースを，共有化メタデータと標準情報検索手順に従う単一データベースであるかのように利用者に見せている。

図3にCIASシステムの検索例を示す。この例ではCIASシステムに対してtitleに「river」という文字列を含んでいるデータの検索を要求し（図3-①），その結果，英国議会資料地図データベースから73件，タミール映画データベースから2件のヒットがあったことを示している（図3-②）。ここで特定のレコードを選択し，更に詳細表示を指示すると，共有化メタデータベースのレコードに保存されている元データベースの該当レコードへのURIを参照し，詳細な書誌情報や画像データなどを表示することができる（図3-③および④）。

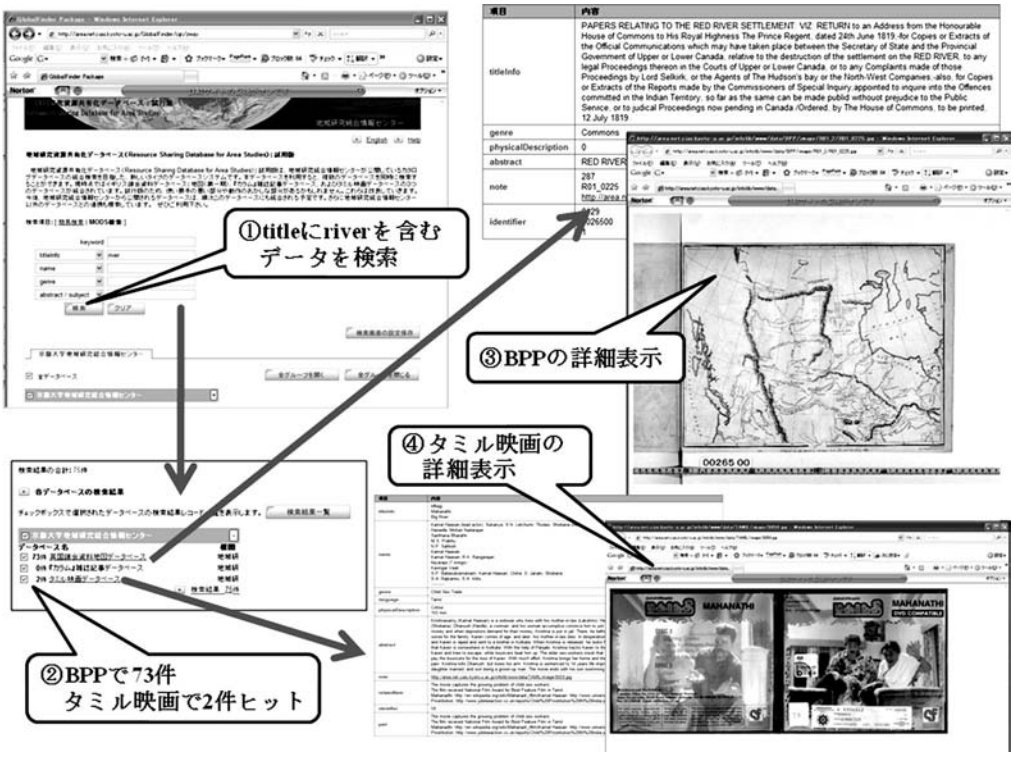

図3　地域研究資源共有化データベースの検索例

III　メタデータの検討

資源共有化システムのデータ共有と検索の能力は，メタデータの設計に大きく依存する。もしメタデータを独自に定義しようとすると，データ項目の選定・粒度・マッピングなど考慮すべき要件が多岐にわたり，多大な時間と労力を必要とする。そこで，資源共有化システムでは，独自にメタデータを設計するのではなく，標準メタデータ[注9)]を基礎とし，必要な修正を加える手法を採用している。

例えば図書館では MARC（MAchine Readable Cataloging）［LC 2004b］や MODS（Metadata Object Description Schema）［LC 2009a］と呼ばれる標準メタデータが普及している。MARC や MODS に従って記述された書誌データは，書誌データベース間で交換可能である。そのため，日本や欧米などでは MARC に基づいた国家規模の書誌ユーティリティが構築されている。アーカイブでは ISAD（G）（General International Standard Archival Description）［ICA 1999］あるいは EAD（Encoded Archival Definition）［LC 2002］と呼ばれる記録史料用の標準メタデータが普及しつつある。博物館では The International Committee for Documentation of the International Council of Museums［2009］などが中心となって博物館情報のメタデータの制定を目指している。

本章では，資源共有化システムのメタデータを設計する上で，これらの標準メタデータをどのように活用すべきか，これまでの研究事例を示しつつ検討する。まずⅢ-1 節において基本的なメタデータである Dublin Core Metadata Element Set の利用例と問題点について述べる。このメタデータは国文学研究資料館の資源共有化システムで利用された。Ⅲ-2 節では Dublin Core Metadata Element Set の問題点を考慮した「拡張された Dublin Core Metadata Element Set」について述べる。このメタデータは NIHU システムで採用されている。

国文学研究資料館や NIHU の資源共有化システムでは，共有化メタデータとして 1 種類の標準メタデータを利用している。これに対して CIAS システムでは，より詳細な検索機能とデジタル資料管理機能の実現を目指し，複数の標準メタデータを組み合わせている。この標準メタデータ組み合わせを，本研究ではメタデータ・スイートと呼んでいる。メタデータ・スイートについてはⅢ-3 節，メタデータ・スイートを構成する主要なメタデータについてはⅢ-4 節で詳述する。

III-1　Dublin Core Metadata Element Set

Dublin Core Metadata Element Set（以下では DCMES）は，多様な情報資源を効率的に発見・交換するための標準メタデータであり，インターネットで最も普及している汎用メタデータの 1 つである。DCMES では基本記述要素と呼ばれる Title，Creator，Subject，

```
〈?xml version＝“1.0” encoding＝“Shift_JIS”?〉
〈record-list〉
  〈dc-record〉
    〈title〉鏑木家〈/title〉
    〈title〉鏑木太郎〈/title〉
    〈creator〉千葉県海上郡海上町史編纂委員会〈/creator〉
    〈subject〉海上町史料所在目録　第三集〈/subject〉
    〈subject〉千葉県海上郡海上町史編纂委員会〈/subject〉
    〈subject〉海上町関係史〈/subject〉
    〈subject〉鏑木家〈/subject〉〈subject〉江戸前〈/subject〉
    〈subject〉下総国　香取郡　鏑木村〈/subject〉
    〈subject〉相給〈/subject〉
    〈description〉海上町関係史料〈/description〉
    〈publisher〉千葉県海上郡海上町史編纂委員会〈/publisher〉
    〈date〉1981〈/date〉
    〈type〉史料所在目録データベース〈/type〉
    〈format〉SGML テキスト〈/format〉
    〈identifier〉1201724〈/identifier〉
    〈source〉nijl.ac.jp〈/source〉
    〈language〉ja〈/language〉
    〈rights〉千葉県海上郡海上町史編纂委員会〈/rights〉
    〈rights〉国文学研究資料館〈/rights〉
  〈/dc-record〉
  …
〈/record-list〉
```

図4　DCMES 基本記述要素によるメタデータの記述例

Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights の15項目のデータ項目を定めている。これはインターネット上の情報資源を，最も一般的と考えられる15種類の特徴で整理しようとするものである。そのため，個別データベースのメタデータと DCMES 間のマッピングは比較的容易であると言われている。

DCMES 基本記述要素のみを利用した資源共有化システム用メタデータの例を図4に示す。これは国文学研究資料館が公開している「史料所在情報・検索」システムの1レコード分を DCMES にマッピングしたものである［原 他 2003: 17–22］。ここで record-list 要素はデータ全体を表す。これは関係データベースのファイルに対応する。dc-record 要素は1つの書誌データを表す。これは関係データベースのレコードに対応し，レコード数だけ dc-record 要素は繰り返される。dc-record 要素の内部にある title 要素や subject 要素は DCMES 基本記述要素を表す。これは関係データベースのフィールドに相当する。実際の書誌データ，例えば title データは，〈title〉と〈/title〉の間に記述されている。

DCMES の問題点の 1 つは基本記述要素の定義が曖昧なことである。例えば，DCMES の定義によれば，Date 要素には資料のライフサイクルに関わる日付・年代・時代などの時間データを記述することになっている。しかし，ライフサイクルに関するどのような時間をどのような記法で Date 要素に記述すべきか，については定義していない。このような定義の曖昧性はマッピングの精度に影響を与える恐れがある。

マッピングの良し悪しが資源共有化システムの検索精度に影響を与えるため，国文学研究資料館の国文学論文目録データを用いて，マッピング精度についての実験を行った［原 他 2005: 31–38］。国文学論文目録データは，日本国内で刊行された雑誌・紀要・単行本などに収められている国文学研究論文の総合目録である。被験者として，国文学研究資料館で目録作成に従事している教職員 6 名（目録系 4 名，アーカイブ系 1 名，情報系 1 名）の参加を仰いだ。実験手順などの詳細は文献に記載している。実験結果の概要であるが，国文学論文目録データから選択されたデータ項目も，マッピング先として選択された DCMES のデータ項目も，被験者により相違が見られた。やはり，DCMES 基本記述要素の定義が曖昧であること，さらに基本記述要素の種類そのものが少なすぎることも原因であるとの結論に至った。

先に「固有メタデータと DCMES 間のマッピングは比較的容易であると言われている」と述べた。しかし，実際の資料を前にしてマッピング規則を定めようとすると，検討しなければならない事項が数多く発生し，想像以上の時間と困難を伴うことが明らかになった［安達 他 2006: 185–214］。経験的には，固有メタデータのデータ項目が多いほど，マッピングの対象とすべきデータ項目の取捨選択とマッピング先の DCMES データ項目の決定にバラツキが大きくなるようであった。

III–2　拡張された Dublin Core Metadata Element Set

資料の内容を DCMES 基本記述要素のみで記述しようとすると，直ちに問題となるのが場所と時間の扱いである。例えば，

- 1 つの資料中に複数の時空間データが記述されていることが多い
- 位置データの種類は地名，住所，緯度・経度など多様である
- 時間データの種類も地方歴，グレゴリオ歴，ユリウス通日など多様である
- 時空間データの記述法も様々である
- データベースをデジタル地名辞書などのアプリケーションプログラムと連携させたい

などの問題や要求に答えることは簡単ではない。これは DCMES 基本記述要素の種類が少ないことに起因する。

DCMES 基本記述要素において，時空間データを対象とする要素は Coverage であり，例えば以下のように記述する。なお以降の XML データでは，簡単のために名前空間[注10)]の接頭辞

を省略している。

〈Coverage〉2008-10-24〈/Coverage〉〈Coverage〉32.30/35.45〈/Coverage〉

〈Coverage〉2009-12-24〈/Coverage〉〈Coverage〉35.45/32.30〈/Coverage〉

この例が2組の時間と位置に関する情報であるらしいとは推察できるが，

a）コンピュータは，どちらが時間データでありどちらが位置データであるか分からない

b）〈Coverage〉32.30/35.45〈/Coverage〉が緯度・経度による位置データであるらしいと推察しても，どちらが緯度でどちらが経度か分からない

c）32.30の単位が度分秒なのか実数であるのか分からない

d）どのような内容の時間と位置なのか分からない

e）時間と位置の関連が分からない（そもそも関連があるのか分からない）

などの問題が発生する。

a）の問題はDCMES基本記述要素の内容を区別する詳細化要素［DCMI 2000］により，ある程度は解決できる。Coverage要素の詳細化要素として，空間データを区別するSpatial要素と，時間データ区別するTemporal要素が定義されている。これらの詳細化要素を使って以下のように記述すれば，時間と空間のデータを区別できる。

〈Temporal〉2008-10-24〈/Temporal〉〈Spatial〉32.30/35.45〈/Spatial〉

〈Temporal〉2009-12-24〈/Temporal〉〈Spatial〉35.45/32.30〈/Spatial〉

b）の問題は地点に関するDCMI Point［DCMI 2006］により解決できる。さらにc）の問題は要素内容の書式を示す符号化形式により解決できる。これらを使って以下のように記述すれば，緯度・経度の区別や単位を明示できる。

〈Point〉

〈north units＝“signed decimal degrees”〉43.9〈/north〉

〈east units＝“signed decimal degrees”〉-80.31〈/east〉

〈/Point〉

しかしd）とe）の問題については，DCMESの詳細化要素の種類にも限界があるので，元データが詳細であればあるほどDCMESのみによる対応は困難となる。そのような場合はDCMESのタグ集合を拡張する。つまり必要なタグ集合を新たに定義し，DCMESと併用することにより，時空間データを詳細に記述する。

拡張されたDCMESを利用した事例としてNIHUシステムにおける時空間メタデータを取り上げる。このメタデータを設計した背景には，歴史研究者からの以下のような要求があった。

- 資料に記載された和暦，時代名あるいは地名で検索したい
- 時間と位置のデータの精度（正確な位置や時間，「～辺り，～頃」など曖昧な位置や時

間，「～カ？」など推測の域をでない位置や時間）を区別したい

- データベースと時空間情報処理アプリケーションプログラムを連携させるため，位置は緯度・経度，時間はグレゴリオ歴に変換しておく必要がある
- データベースからデジタル地名辞書を利用できるように，地名は適切な単位（村，旧国など）に分解しておく必要がある
- データベースから暦日テーブルを利用できるように，和暦は元号と年・月・日に分解しておく必要がある

DCMES の詳細化要素や符号化形式だけでは，これらの要求に対応できないため，NIHU システムの時空間メタデータでは，Coverage 要素の内容を詳細に記述するためのタグ集合を独自に定義した（付録１）。DCMES と独自に定義したタグ集合は区別する必要がある。そこで，DCMES については名前空間

xmlns:dc＝“http://purl.org/dc/elements/1.1/”

に対して「dc」という接頭辞を割り当てた。また独自に定義したタグ集合には名前空間

xmlns:ts＝“http://www.nihu.jp/ResourceSharing/GeoTemporal”

を定義し，それに「ts」という接頭辞を割り当てた。

付録１のメタデータ定義に従うと，時間（ts: Date 要素）は開始時間（ts: From 要素）と終了時間（ts: To 要素）で区切られた範囲として記述される。１時点を表す場合，ts: From 要素と ts: To 要素の内容は同じになる。時間データは，資料に記載されている時間をそのまま記述する部分（ts: DescDate 要素）と，グレゴリオ歴に正規化された時間を記述する部分（ts: NormDate 要素）から構成されている。データベースから暦日テーブルを参照できるように，ts: description 要素に記述された和暦データは，ts: regionalCal 要素の内部で元号（ts: gengo 要素），年（ts: year 要素），月（ts: month 要素），日（ts: day 要素）に分解して記述できるようになっている。

場所（ts: Place 要素）は 北西端（〈ts: NWpoint〉）と南東端（〈ts: SEpoint〉）で示された矩形範囲として記述される。１地点を表す場合，ts: NWpoint 要素と ts: SEpoint 要素の内容は同じになる。地名データは，資料に記載されている地名をそのまま記述する部分（ts: DescPlace 要素）と，現在の地名および緯度・経度に正規化された位置を記述する部分（ts: NormPlace 要素）から構成されている。データベースからデジタル地名辞書を参照できるように，ts: description 要素に記述された地名データは，ts: address 要素の内部で町名（ts: town 要素），郡市（ts: couty 要素），県・旧国（ts: prefecture 要素）などに分解して記述できるようになっている。

データの精度を表現するために，ts: desc_precise という属性[注11)]を定義している。位置あるいは時間データが正確であれば，属性値を「accurate」とする。これが ts: desc_precise 属

```
〈?xml version＝“1.0”〉
〈ts:RecordData
  xmlns:dc＝“http://purl.org/dc/elements/1.1/”
  xmlna:ts＝“http://www.nihu.jp/ResourceSharing/GeoTemporal”〉
  〈ts:DC〉
    〈dc:Title〉末醤　2升（2升）〈/dc:Title〉
    〈dc:Description〉［単価］新銭（10文/升）〈/dc:Description〉
    〈dc:Description〉［備考］奉写一切経料。〈/dc:Description〉
    〈dc:Date〉宝亀2（771）年8月22日〈/dc:Date〉
    〈dc:Type〉［品目］食料（調味料）〈/dc:Type〉
    〈dc:Relation〉［史料］奉写一切経料銭用帳/続々修2-8/大日本古文書17-323〈/dc:Relation〉
    〈dc:Coverage〉
      〈ts:Container〉
        〈ts:Date〉
          〈ts:From〉
            〈ts:DescDate〉
              〈ts:description〉宝亀2（771）年8月22日〈/ts:description〉
              〈ts:regionalCal〉
                〈ts:gengo〉宝亀〈/ts:nengo〉
                〈ts:year〉2〈/ts:year〉
                〈ts:month〉8〈/ts:month〉
                〈ts:day〉22〈ts:day〉
              〈/ts:regionalCal〉
            〈/tsDescDate〉
            〈ts:NormDate ts:norm_format＝“W3C”〉0771-10-04T00:00:00＋09:00〈/ts:NormDate〉
          〈/ts:From〉
          〈ts:To〉
            〈ts:DescDate〉
              〈ts:Description〉宝亀2（771）年8月22日〈/tsDescription〉
              〈ts:regionalCal〉
                〈ts:gengo〉宝亀〈/ts:nengo〉
                〈ts:year〉2〈/ts:year〉
                〈ts:month〉8〈/ts:month〉
                〈ts:day〉22〈ts:day〉
              〈/ts:regionalCal〉
            〈/tsDescDate〉
            〈ts:NormDate ts:norm_format＝“W3C”〉0771-10-04T00:00:00＋09:00〈/ts:NormDate〉
          〈/ts:To〉
        〈/ts:Date〉
        〈ts:Place〉
          〈ts:DescPlace〉
            〈ts:description〉［地域］大和国〈/ts:description〉
            〈ts:address〉
              〈ts:prefecture〉大和国〈/ts:prefecture〉
            〈/ts:address〉
```

図5　拡張されたDCMESによるメタデータの記述例

```
            〈ts:DescPlace〉
            〈ts:NormPlace〉
              〈ts:coodinate〉
                〈ts:NWpoint〉
                  〈ts:x34.448611〈/ts:x〉
                  〈ts:y〉135.799444〈/ts:y〉
                〈/ts:NWpoint〉
                〈ts:SEpoint〉
                  〈ts:x〉34.448611〈/ts:x〉
                  〈ts:y〉135.799444〈/ts:y〉
                〈/ts:SEpoint〉
              〈/ts:coodinate〉
            〈/ts:NormPlace〉
          〈/ts:Place〉
        〈/ts:Container〉
      〈/dc:Coverage〉
    〈/ts:DC〉
  〈/ts:RecordData〉
```

図 5 ―続き

性のデフォルト値である。「～辺り，～頃」など，データが曖昧である場合は，属性値を「about」とする。さらに「～カ？」など，データが推察の域を出ないような場合は，属性値を「maybe」とする。この定義に従ったメタデータの記述例を図 5 に示す。

図 5 では，１つの事象に対して和暦と西暦，旧地名と緯度・経度など，複数の時空間データが与えられている。この例が示すように，DCMES 基本記述要素のみの場合に比べると，時空間データはかなり詳細に記述できるようになった。ただし空間データは矩形のみであり，河川や湖などの線分や多角形は記述できないという問題点がある。これについてはメタデータの拡張を考えている。またデータ構造が複雑であり，手作業でデータを記述することは困難であるとの指摘もある。これについてはテンプレートや簡易エディタなど方策を考える必要がある。

III-3　メタデータ・スイート（Metadata Suites）

これまでの検討から，拡張された DCMES は共有化メタデータとして有用であることが示された。それでもデータ項目数が制限されているため，書誌データなどを拡張された DCMES にマッピングすると情報の劣化は避けられない。そこで，情報の劣化を軽減しつつ多様なデータの共有化を実現させるために，資源共有化システムモデルの見直しに着手した。以下では地域研のデータベースを例として検討を進める。

地域研が公開中あるいは公開準備中のデータベースの一覧を示す。本稿執筆時点において①から③のデータが CIAS システムで共有化されている。

①英国議会資料地図データベース（公開中）: 固有メタデータ＋地図画像

②カラム雑誌記事データベース（公開中）: 固有メタデータ＋記事画像

③タミール映画データベース（公開中）: 固有メタデータ＋ジャケット画像（映像は準備中）

④三印法典データベース（公開中）: 全文データベース（固有 XML マークアップ）

⑤地域研究資源アーカイブ（準備中）: 標準メタデータ（EAD，MODS，METS）＋写真画像

⑥マレーシア映画データベース（準備中）: 固有メタデータ＋ジャケット画像＋映像

⑦イスラム雑誌記事データベース（準備中）: 固有メタデータ＋記事画像

⑧トルキスタン集成データベース（準備中）: 固有メタデータ（画像は準備中）

⑨ポスト社会主義諸国選挙・政党データベース（準備中）: ファクトデータ（固有形式）

これらのデータベースは以下の3つのタイプに分類できる。

タイプⅠ）本質的に固有メタデータであるデータベース: ④（全文データベース）と⑨（表形式のデータ）が該当する。

タイプⅡ）標準メタデータに準拠可能であるが，既に固有メタデータにより構築されているデータベース: ①，②，③，⑥，⑦，⑧が該当する。基本的に書誌データベースである。

タイプⅢ）標準メタデータに準拠しているデータベース: ⑤が該当する。

このように多様な資料の内容を詳細に記述し，さらに情報の劣化を押さえつつ多様なデータの共有化を実現する方法として，メタデータの使い分けを試みている。もし書誌データベース同士のように同質のデータを共有化するのであれば，DCMES のような汎用メタデータではなく書誌専用のメタデータを利用した方が，情報の劣化を防ぐことができる。一方，図書館と博物館のように異質なデータを共有化するのであれば，情報の劣化を許容して汎用メタデータを利用する。このように共有化するデータの性質に合わせてメタデータを使い分ける。メタデータ・スイートは，このような発想から作られたモデルであり，下層から順に Database Layer，Standard Layer，Sharing Layer，Application Layer の4層から構成されている（図6）。

メタデータ・スイートの最下層を Database Layer と呼ぶ。これは資源共有化システムモデル（図2）の Database Layer と同じく，非標準メタデータの層である。ここには固有メタデータに基づいたタイプⅠとタイプⅡのデータベース，およびタイプⅢのデータベースが対象とする写真などのデジタルデータファイルが該当する。データベースシステムの効率的な管理・運用および共有化を考えると，固有メタデータの多用は望ましくない。しかしメタデータの標準化が遅れている分野も多くあること，標準化を強制すると新しい可能性の芽を摘んでしまう恐れがあり自由度を大きくしたいという要望があることなどを考慮すると，非標準メタ

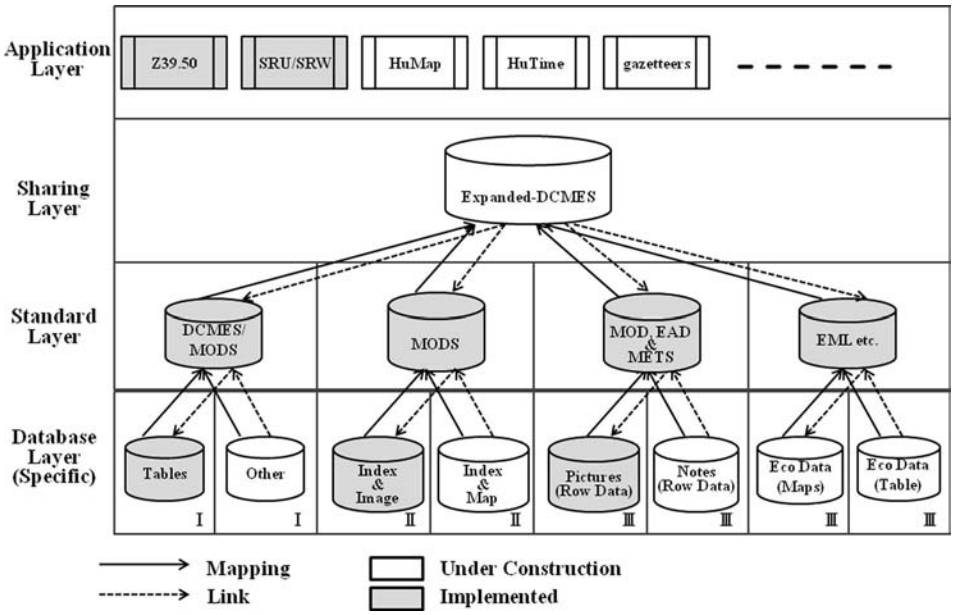

図 6　メタデータ・スイートのモデル

データを対象とした Database Layer は必要である。

Database Layer の上位層を Standard Layer と呼ぶ。これは資源共有化システムモデルの Database Layer と Sharing Layer の中間に位置する層である。ここには分野ごとの標準メタデータが用意されている。例えば本・論文・地図などの資料を 1 点ごとの書誌データとして記述するなら MARC などを利用する。Database Layer に該当する個別データベースは，Standard Layer 内の適切な標準メタデータにマッピングされる。つまり Standard Layer において，分野ごとの共有化を実現する。特定分野の標準メタデータのデータ項目には該当分野の特徴が反映されており，データ項目数も DCMES に比べて豊富である。DCMES よりも精度の高いマッピングが可能であり，情報や検索精度の劣化を軽減できる。さらに特定分野内におけるデータベース横断検索の方が，分野横断的なデータベース検索よりも需要が多い。このように，Standard Layer における分野ごとの資源共有化は合理的であると言える。

Standard Layer の上位層を Sharing Layer と呼ぶ。これは資源共有化システムモデルの Sharing Layer と同じ機能の層である。DCMES のような汎用メタデータを利用する。Standard Layer が分野ごとの標準メタデータを利用して精度の高い共有化を実現する層であるのに対して，Sharing Layer では分野横断的な共有化を実現する。Standard Layer より幅広い検索は可能であるが，検索精度は低下する。なお本稿執筆時点において，CIAS システムでは Sharing Layer を実装していない。

メタデータ・スイートの最上層を Application Layer と呼ぶ。これも資源共有化システムモ

デルの Application Layer と同じ機能の層である。メタデータの層ではなく，資源共有化システムを外部から利用するための機能を集めた部分である。他の資源共有化システムと連携するためのゲートウェイ機能や，HuMap あるいは HuTime などの時空間情報処理アプリケーションプログラムと連携するための API（Application Program Interface）[注12)] などを想定している。本稿執筆時点の CIAS システムでは，Z39.50 と SRU/SRW のみが実装されている。

メタデータ・スイートのメタデータは，下層ほど分野特異性が強く，階層を上るにつれて一般的つまり共有化の適用範囲が広くなる。メタデータ・スイートは，資料の特性に応じて各層から適切なメタデータを選択するためのガイドラインと言うこともできる。

III-4　メタデータ・スイートの主要なメタデータ

メタデータ・スイートモデルに特徴的な層は Standard Layer である。CIAS システムの Standard Layer で用意されている主要なメタデータは，DCMES，MODS，EAD および METS である。

CIAS システムでは，タイプⅠのデータを共有化するメタデータとして DCMES あるいは MODS[注13)] を採用している。なお本稿執筆時点では MODS のみを利用している。以下に述べるタイプⅡとⅢのデータを共有化するために MODS を実装しているためである。

前節のデータベース一覧から分かるように，地域研データベースの多くは雑誌・映画・地図などの書誌的データである。このようなタイプⅡおよびⅢの書誌的データを共有化するメタデータとして，CIAS システムでは MODS を採用している。MARC よりも簡便でありながら DCMES よりも詳細な記述が可能な MODS の特徴を評価したためである。CIAS システムでは，個別書誌データベースのデータ項目を MODS の適切なデータ項目にマッピングして MODS データベースを構築し，これを Standard Layer の共有化メタデータベースとしている。図3に示した CIAS システムの検索例は，このレベルにおける共有化である。

タイプⅡおよびⅢのアーカイブデータを共有化するメタデータとして CIAS システムでは EAD と METS を採用している。以下では EAD と METS について詳述する。

［EAD］

フィールドノート・写真・地図などは地域研究における貴重な資料であるが，研究者が退職すると廃棄・散逸する傾向にあり，これらの資料の保存と公開は焦眉の問題となっている。そこで，地域研では CIAS システム開発の一環として，写真を中心としたコレクションのデータベース化に着手した。

このようなコレクションのデータベース化においては，資料1点ごとの目録化だけでは不十分であり，資料の出所・来歴など資料間の依存関係も記述する必要がある。CIAS システムで

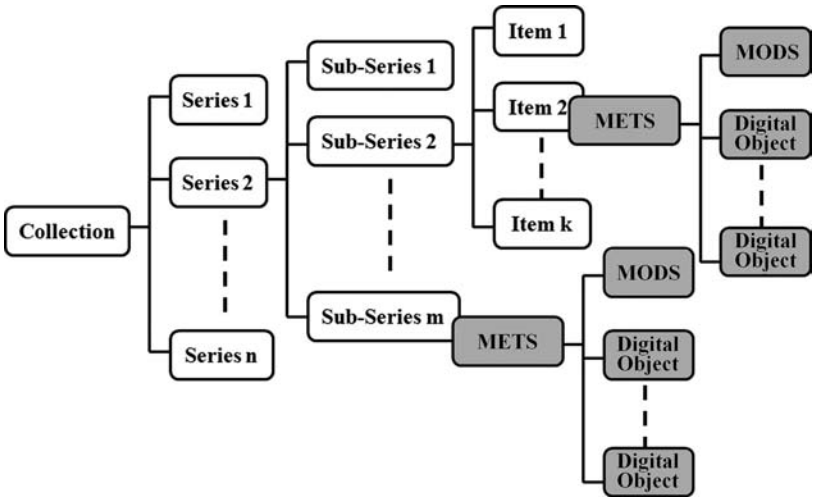

図 7　EAD によるアーカイブの階層モデル

は，アーカイブにおける標準メタデータである EAD を導入し，資料間の依存関係を階層的に記述している[注14)]。

EAD による資料の階層構造を図 7 に例示する。この図では，資料を Collection，Series，Sub-Series および Item の 4 階層で構造化している。ここで Collection はある資料群を表し，その Collection は 1 つ以上の Series から，Series は 1 つ以上の Sub-Series から，さらに Sub-Series は 1 つ以上の Item から構成されている。階層を下るにつれて資料の単位は小さくなり，Item が 1 点ごとあるいは少数の資料の塊となる。例えば石井米雄先生撮影の東南アジア写真資料の場合，Collectionは「石井米雄氏写真資料目録」であり，Series はスライドファイルなど納めた 1 個ずつの段ボール箱，Sub-Series は段ボール箱の中にある 1 冊ずつのスライドファイルなどである。それ以下のレベルについては整理中であり，現時点でスライドファイルの 1 ページごとが Item となっている（図 7 中の METS については後述する）。

この EAD データを XML で記述した例を図 8 に示す。[5)] EAD の基本構造は eadheader（EAD ヘッダ，図 8-Ⅰ），frontmatter（前付け，図 8-Ⅱ）および archdesc（記録史料記述，図 8-Ⅲ）の 3 つの要素である。eadheader 要素には，この記録自体の書誌記述，frontmatter 要素には序文・献辞・凡例・この記録データの使用法などを記述する。実際の検索データは archdesc 要素内に階層的に記述される。

図 8 の例では archdesc 要素が Collection に対応する。その内部の C01 要素（図中のⅣ）が Series で，1 個ずつの段ボール箱に対応する。さらにその内部の C02 要素（図中のⅤ）が

5）図 8 および図10中のローマ数字は説明のために追加したもので，XML データではない。

```
〈ead audience＝“internal”〉
（Ⅰ）
〈eadheader langencoding＝“iso639-2b” …〉
  〈eadid url＝…〈/eadid〉
  〈filedesc〉
    〈titlestmt〉
      〈titleproper encodinganalog＝“Title”〉石井米雄氏写真資料目録…〈date〉1957–1968〈/date〉
      〈/titleproper〉
      〈titleproper type＝“filing”〉石井米雄氏写真資料（仮）Ishii (Yoneo) Photographs〈/titleproper〉
      〈author encodinganalog＝“Creator”〉記述・編集 Finding Aid Written by: 五島敏芳.
      〈/author〉
    〈/titlestmt〉
    〈publicationstmt〉
      〈publisher encodinganalog＝“Publisher”〉Library of Center for Integrated…〈/publisher〉
      …
〈/eadheader〉
（Ⅱ）
〈frontmatter〉
  …
  〈head〉まえがき Preface 〈/head〉
  〈p〉このファイルは，標記資料の電子的検索手段である．その記述の範囲・水準・規則については，凡例か本文のなかで説明される．〈/p〉
  〈p〉〈head〉まえがき Preface〈/head〉
    …
  〈p〉1. この目録は，「石井米雄氏写真資料」（仮）の資料記述を収録した．〈/p〉
  〈p〉2. 資料記述は，DACS，Describing Archives: A Content Standard (Society of American…
…
〈/frontmatter〉
（Ⅲ）
〈archdesc relatedencoding＝“Dublin Core” level＝“collection” type＝“inventory”〉
  〈did〉
    〈head〉[資料の概観 Collection Summary]〈/head〉
      〈repository encodinganalog＝“Publisher” label＝“収蔵 Repository:”〉
        〈corpname〉京都大学地域研究統合情報センター…〈/corpname〉
        …
      〈/repository〉
      〈origination label＝“出所 Origination/作成 Creators:”〉
        〈name encodinganalog＝“Creator”〉石井米雄〈/name〉
      〈/origination〉
      …
  〈/did〉
    …
〈dsc type＝“combined”〉
〈head/〉
（Ⅳ）
```

図8　EADによる資料の依存関係の記述例

```
  〈c01 level＝“series” id＝“cias-2006001.01”〉
    〈did〉
      〈unitid encodinganalog＝“Identifier” label＝“資料記号 Reference code:”〉2006001/1〈/unitid〉
        …
      〈extent unit＝“件items” type＝“数量 Extent:”〉8〈/extent〉8 冊．（内訳）小豆色 7 冊，紺色 1
      冊．〈/physdesc〉
    …
    〈/did〉
      …
（V）
    〈c02 level＝“subseries” id＝“cias-2006001.0201”〉
      〈did〉
        〈unitid encodinganalog＝“Identifier” label＝“資料記号 Reference code:”〉2006001/2-1〈/unitid〉
        〈unittitle encodinganalog＝“Title” id＝“cias-2006001.0201” label＝“標題・年代 Title and
        Date:”〉
          [SLIDE FILE] 1970s 貝葉調査　タイ．
          〈unitdate encodinganalog＝“Coverage (Temporal)” normal＝“1970”〉（昭和45年［1970］）．
          〈/unitdate〉
        〈/unittitle〉
        〈physdesc label＝“形態等状態 Physical Descriptions:”〉
          〈extent unit＝“件 items” type＝“数量 Extent:”〉1〈/extent〉
          スライドファイル 1 ケース．（内訳）: スライドシート 9 枚; 計約180コマ（4 行× 5 コマ/
          1 シート）．
        〈/physdesc〉
        〈langmaterial encodinganalog＝“Language”〉
          〈language〉日本語 Japanese，英語 English.〈/language〉
        〈/langmaterial〉
      〈/did〉
      〈scopecontent〉
        〈head〉範囲と内容 Scope and Content 〈/head〉
          〈p〉表紙印刷:「FUJICOLOR SLIDE FILE」．内容: 北タイ，チェンマイあたりヵ．…〈/p〉
      〈/scopecontent〉
      〈controlaccess/〉
      〈phystech〉
        〈head〉物的特徴 Physical Characteristics/技術要件 Technical Requirements 〈/head〉
        背表紙下部附箋「(26)　179枚」．標題は，ファイル状のケースの背表紙より採取．［　］内
        は，あらかじめ印刷されていた部分．ほか手書き部分は，油性マジックにより記載
      〈/phystech〉
    〈/c02〉
  …
  〈/c01〉
〈/dsc〉
〈/archdesc〉
〈/ead〉
```

図 8―続き

Sub-Series で，1冊ずつのスライドファイルに対応する。この例では C01 要素と C02 要素は省略されて1つしか見えないが，実際には箱数だけ，また1個の箱の中にあるスライドファイルの冊数分だけ繰り返し記述される。

［METS］

EAD では写真などのデジタルコンテンツを Item に割り当てるのが基本であるが，Series や Sub-Series などに割り当てても構わない。Sries，Sub-Series あるいは Item などにデジタルコンテンツを割り当てる場合，①デジタルコンテンツ自身に関するデータ，②デジタルコンテンツ1点ごとの書誌データ，③資料間の依存関係を示すデータ，④デジタルコンテンツに関する技術的なデータ，[注15)] の4種類の情報をまとめてパッケージとして記述する必要がある。

CIAS システムでは，METS（Metadata Encoding & Transmission Standard）［LC 2009b］を採用し，デジタルコンテンツの関連情報を記述している。図7の網掛け部分が METS パッケージであり，EAD との関係を示している。

METS パッケージの構造を図9に示す。METS パッケージ内にはデジタルコンテンツ情報と保存情報が含まれている。CIAS システムの場合，デジタルコンテンツ情報は上記①のデジタルコンテンツ自身に関するデータに対応し，その実体はデジタルデータファイルへのリンク情報である。保存情報はデジタルコンテンツに関するメタ情報群で，上記の②から④に対応する。②のデジタルコンテンツ1点ごとの書誌データは，デジタルコンテンツ自身の書誌データと，デジタルコンテンツが付与されている EAD 要素（Series，Sub-Series，Item など）に関するデータの2つから構成されている。②のデジタルコンテンツ自身の書誌データは MODS で記述している。③の資料間の依存関係情報に対応するデータは，デジタルコンテンツが付与されている EAD 要素までの階層データである。上記④のデジタルコンテンツに関する技術データについては検討中であり，本稿執筆時点において METS パッケージには収容されていない。

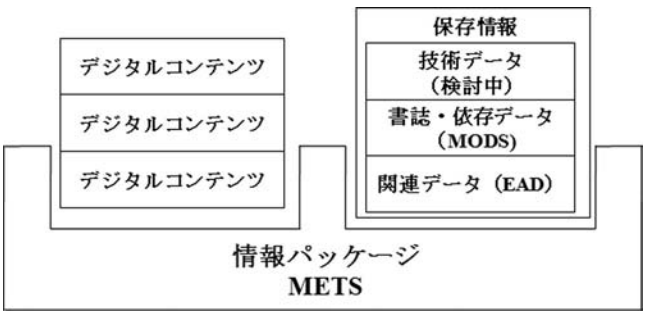

図9　METS パッケージの構造

```
〈?xml version＝“1.0” encoding＝“UTF-8”?〉
〈mets:mets　※名前空間の記述は省略〉
（Ⅵ）
〈mets:metsHdr CREATEDATE＝“2008-10-18” LASTMODDATE＝“2008-10-20”〉
  〈mets:agent ROLE＝“CREATOR” TYPE＝“ORGANIZATION”〉
    〈mets:name〉Library of Center for Integrated Area Studies, Kyoto University〈/mets:name〉
    …
〈/mets:metsHdr〉

（Ⅶ）
〈!―EAD データを分解，記述レベルごとの記述単位データを抽出．（資料管理者側）―〉
〈mets:dmdSec ID＝“DMD1”〉
  〈mets:mdWrap MDTYPE＝“OTHER” OTHERMDTYPE＝“subEAD”〉
    〈mets:xmlData〉
      〈subEAD:subEAD〉
        〈subEAD:c02 level＝“subseries” id＝“cias-2006001.0201”〉
    …
〈/mets:dmdSec〉

（Ⅷ）
〈!―MODS データで記述．（利用者/研究者側）―〉
〈mets:dmdSec ID＝“DMD2”〉
〈mets:mdWrap MDTYPE＝“MODS”〉
  〈mets:xmlData〉
    〈mods:mods〉
      〈mods:titleInfo〉
        〈mods:title〉[SLIDE FILE] 1970s 貝葉調査　タイ〈/mods:title〉
…
〈/mets:dmdSec〉
〈!―権利関係管理メタデータ―〉
〈mets:amdSec〉…
〈/mets:amdSec〉

（Ⅸ）
〈!―デジタルデータへのリンク情報―〉
〈mets:fileSec〉
  〈mets:fileGrp〉
    〈mets:file ID＝’cias-2006001.0201_001’〉
      〈mets:FLocat LOCTYPE＝’OTHER’ xlink:href＝’07box2-01/07box2-01_001.JPG’/〉
    〈/mets:file〉
    …
  〈/mets:fileGrp〉
〈/mets:fileSec〉

（Ⅹ）
```

図10　METS によるパッケージ情報の記述例

```
〈!―つぎの structMap は，資料群内の階層構造を示し，この METS 文書の対象となる資料記述の記述レベルにいたるまでの各記述レベルすべての標題のみをあげる―〉
〈mets:structMap LABEL＝“Archival structure” TYPE＝“logical”〉
  〈mets:div〉
    〈mets:div ID＝“CIAS-KYOTO-U” TYPE＝“repository”〉
      〈mets:div LABEL＝“title”〉Library of Center for Integrated Area Studies, Kyoto University
      〈/mets:div〉
    〈/mets:div〉
    〈mets:div ID＝“cias-2006001” TYPE＝“collection”〉
      〈mets:div LABEL＝“title”〉石井米雄氏写真資料 Ishii (Yoneo) Photographs〈/mets:div〉
    〈/mets:div〉
    〈mets:div ID＝“cias-2006001.02” TYPE＝“series”〉
      〈mets:div LABEL＝“title”〉(箱 2: SLIDE FILE, …〈/mets:div〉
    〈/mets:div〉
    〈mets:div ID＝“cias-2006001.0201” TYPE＝“subseries”〉
      〈mets:div LABEL＝“title”〉[SLIDE FILE] 1970s 貝葉調査　タイ. 〈/mets:div〉
    〈/mets:div〉
  〈/mets:div〉
〈/mets:structMap〉
〈/mets:mets〉
```

図10―続き

図10に CIAS システムにおける METS データの例を示す。これは図8の EAD データ（図8-Ⅴ）部分

〈c02 level＝“subseries”id＝“cias-2006001.0201”〉

に関連するパッケージ情報である。METS は metsHdr（METS header: ヘッダ，図10-Ⅵ），dmdSec（descriptive metadata: 記述的メタデータ，図10-ⅦおよびⅧ），fileSec（file section: ファイルセクション，図10-Ⅸ）および structMap（structural map: 構造マップ，図10-Ⅹ）などから構成されている。

metsHdrr 要素には，この記録自体の書誌データを記述する。例えば図10-Ⅵには，保存機関名が地域研であることなどが記載されている。

dmdSec 要素は上記②に対応しており，デジタルコンテンツに関する書誌情報を記述する。図10-Ⅶには，このデジタルコンテンツの付与されている EAD 要素が id＝“cias-2006001.0201” で識別される Sub-Series 要素であることなどが記載されている。また図10-Ⅷには，このコンテンツのタイトルが「[SLIDE FILE] 1970s 貝葉調査　タイ」であるなど，デジタルコンテンツ自身に関する書誌データが MODS で記載されている。

fileSec要素は上記①に対応しており，デジタルコンテンツファイルへのリンク情報を記述する。図10-Ⅸには，デジタルコンテンツファイルの名前が “07box2-01/07box2-01_001.JPG” であることなどが記載されている。

図11　地域研究資源アーカイブの検索例

structMap要素に上記③に対応しており，デジタルコンテンツが付与されている EAD 要素までの階層データを記述する。図10-Ｘには，デジタルコンテンツ石井米雄氏写真資料（Collection）の第２箱目（Series）の「[SLIDE FILE] 1970s 貝葉調査 タイ」というタイトルのスライドファイル（Sub-Series）に含まれていることなどが記載されている。

Standard Layer において EAD，MODS および METS を組み合わせた CIAS システムのアーカイブ検索例を図11に示す。この例からも明らかなように，DCMES よりも詳細な内容記述が可能となっている。

IV　資源共有化システムの今後の展開

これまでの研究・開発により，地域研究における資源共有化システムの基盤は整備された。CIAS システム以前に開発された NIHU システムなどに比べると，CIAS システムは共有化に伴う情報や検索精度の劣化を軽減できる，写真などのコレクション資料を扱える，などの優位性を実現している。以下では CIAS システムの課題と今後の展開について検討する。

まずMODSやEADなどのメタデータ構造が複雑であること，さらにXMLによる記述を前提としているため，地域研究者には敷居が高いという問題が指摘されている。複雑なメタデータ構造への対策として，best practice guideの作成を進めている。これは，各メタデータ項目に対して，「DCMESのTitleには副題も含める」のように記述すべき内容を明確化し，「DCMESのSubjectは米国議会図書館のサブジェクトスキーマに従って記述する」のように記述法を定義するとともに，事例を示したものである。データ入力については，アーキビストなどの専門家用には国文学研究資料館やUniversity of California Berkeleyの図書館が開発したEAD Xpress［Conkin 2006］などのメタデータエディタ，研究者用にはMicrosoft Excelなどを利用したテンプレートの利用を考えている。

資源共有化システムの実現により多様なデータの共有化は容易となったが，その一方で資源共有化システムは研究支援あるいは知識発見のツールになり得ていないという意見も聞かれる。たしかに，従来のデータベースの主たる目的はデジタル資料の保存と書誌データの管理であり，メタデータの記述は表層的なレベルにとどまっている。内容に関するデータ項目も多少はあるが，これだけでは研究に供することはできない。この問題については，地域情報学モデル（図1）の知識生成で述べた，HuMapやHuTimeなどの時空間情報処理アプリケーションプログラムが解決のカギとして期待される。現状の資源共有化システムは語彙による情報検索システムにすぎないが，HuMapやHuTimeと組み合わせれば，特定の地域・時代に関するデータを検索する，データ発生地点の分布を様々な地図と重ね合わせて考察する，一連の事象を様々な年表と重ね合わせて考察する，統計処理を行う，語彙だけでは見えなかった関連を時空間の視点から発見する，ことなどが可能となる。つまり，時空間情報処理アプリケーションプログラムと組み合わせることにより，資源共有化システムは強力な研究支援ツールとなる可能性を秘めている。これが現在進行中の資源共有化システムにおける研究・開発テーマである。

時空間情報処理アプリケーションプログラムが真価を発揮する前提として，資料内容に関する時間・空間・概念・事件・人物などの主題を詳細に記述する必要がある。これまで，デジタル地名辞書や暦日テーブルなど時空間に関する意味的情報の組織化には努めてきたが，概念・事件・人物などの主題については手をつけることができなかった。このような主題に関する語彙の収集と整理が今後の課題となろう。ただし，単なる辞書やシソーラスでは不十分であり，オントロジーの導入，つまりコンピュータが概念を処理できるような形式で語彙を体系化する必要がある。オントロジーの導入により，単純な語彙検索では不可能であった様々な関連を発見できる可能性が開けてくる。

ところで，主題間の関連づけを仮に「知識」とするならば，デジタル地名辞書や暦日テーブルを構築し，これらの主題をTopic Maps［ISO 1999］[注16)]などのオントロジー技術を利用し

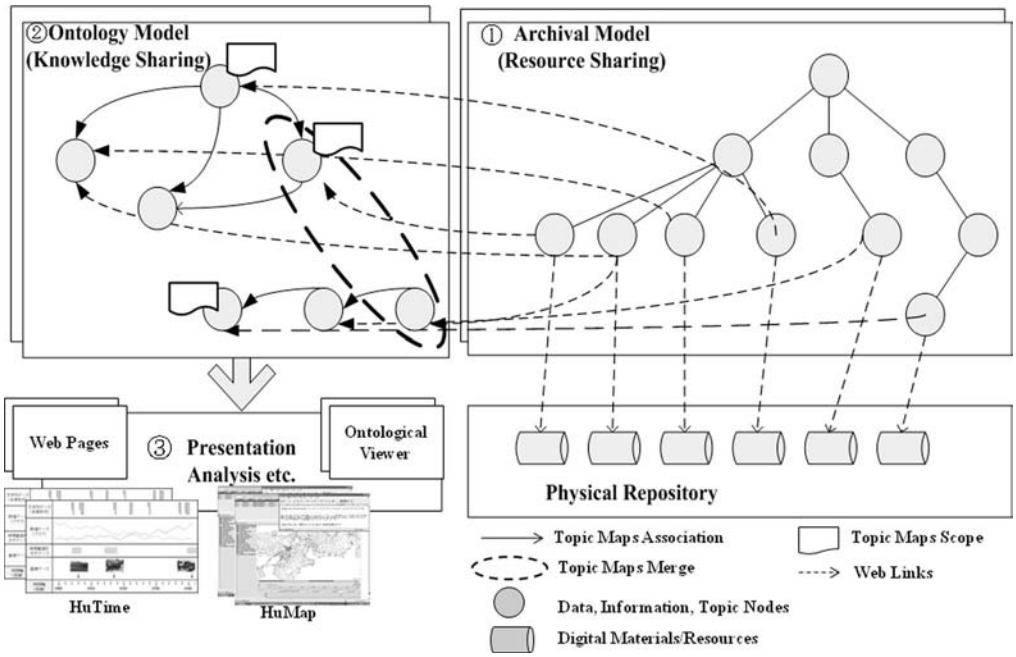

図12　知識共有化モデル

て関連付ける試みは，知識利用の初歩的な事例と考えられる。主題間の関連付けは，それを行った研究者の視点・目的・経験・研究分野などに依存するので個別的である。つまり同じ対象に対して複数の知識が存在しうる。もし多様な知識を形式的に記述・蓄積できれば，それらの知識を組み合わせることにより，分野横断的な知識の発見も可能となろう。地域研究の特徴は，ある地域を多面的に分析することであるが，個人研究者の関与できる範囲は限られている。各研究者の資料に対する見方・関連づけ・処理法などを知識として集積できれば，同一地域における複眼的な視点，他地域との相関的研究を実現できるかもしれない。本研究では，これを知識共有化と呼んでいる。モデルを図12に示す。

このモデルの特徴は，資料の情報を，モノの管理と知識の管理の部分に明確に分離している点にある。書誌などの表層的な情報の管理はデジタルアーキビストなどが担当し，これまで述べた資源共有化システムを利用する。知識は個別的であるので，モノの管理から独立して研究者が担当し，オントロジーを利用する。Topic Maps は主題を形式的に関連づけるツールの候補であり，地域研では写真コレクションを素材に，Topic Maps の適用を試みている。

資源共有化システムの研究・開発は新世紀の開始とともに始まり，ほぼ８年が経過した。国文学研究資料館の内部横断検索システムとして細々と始まった研究も，NIHU システムでは機関間の横断検索システムに発展し，更に横断検索の対象を機構外にまで広げようとしている。一方，CIAS システムは，新しい情報技術を取り入れて，より機能の高いシステムへの脱皮を

図っている。その成果は京都大学研究資源アーカイブへも適用されている。これらの過程で，データ・情報・知識を相互に「連携させる」あるいは「繋げる」ことの意義が浸透しつつあり，デジタル地名辞書，暦日テーブルからオントロジーへと展開している。資源共有化システムは研究途上であり，多様な資料を対象として試行錯誤を繰り返している。図2，図6，図12が示すように，関連するモデルもややまとまりを欠いている。しかし，これは次第に収斂していくものと考えている。資源共有化システムの今後は，資料管理ツールから知識発見ツールへの機能拡張であり，そこではオントロジーが主要な役割を担うものと考えている。

謝　　辞

本稿の執筆にあたり H-GIS（Humanities GIS）研究会メンバーの多大な貢献があった。特に HuTime については国立民族学博物館の久保正敏教授および総合地球環境学研究所の関野樹准教授，デジタル地名辞書については大阪国際大学の桶谷猪久夫教授，暦日テーブルについては国文学研究資料館の相田満助教，EAD・METS データの作成に当たっては京都大学総合博物館の五島敏芳講師，地域情報学の概念形成については京都大学東南アジア研究所の柴山守教授および神奈川大学の貴志俊彦教授のお世話になった。ここに心から謝意を表したい。

参 考 文 献

安達文夫; 鈴木卓治; 小島道裕; 高橋一樹．2006.「情報資源共有化のための博物館資料――データベースのマッピングとその評価」『国立歴史民俗博物館研究報告』125: 185–214．国立歴史民俗博物館.

Alexandria Digital Library．2004. *Alexandria Digital Library Project Gazetteer Development.* http://www.alexandria.ucsb.edu/gazetteer/.

ANSI/NISO．1995. *ANSI/NISO Z39.50-1995 Information Retrieval(Z39.50) Application Service Definition and Protocol Specification.*

地域研究統合情報センター．2009.「地域研究資源共有化データベース（試行版）」．http://www.cias.kyoto-u.ac.jp/index.php/news_detail/id/174.

Conkin, Michael．2006. *EAD XPress.* http://www.lib.berkeley.edu/digicoll/bestpractices/ead_tools.html.

大学共同利用機関法人人間文化研究機構．2009.「研究資源共有化システム」．http://www.nihu.jp/kyoyuka/database.html.

Dublin Core Metadata Initiative (DCMI)．2000. *Dublin Core Qualifiers.* http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/.

――――．2006. *DCMI Point Encoding Scheme: A Point Location in Space, and Methods for Encoding This in a Text String.* http://dublincore.org/documents/dcmi-point/.

――――．2008. *Dublin Core Metadata Element Set, Version 1.1.* http://dublincore.org/documents/dces/.

The Getty．2009. Getty Thesaurus of Geographic Names Online. http://www.getty.edu/research/conducting_research/vocabularies/tgn/.

原　正一郎．2002a.「国文学支援のための SGML/XML データシステム」『情報知識学会論文誌』11(4): 17–35．情報知識学会.

――――．2002b.「Z39.50 とメタデータによる研究機関間連携」『情報処理』43(9): 968–974．情報知識学会.

――――．2007.「人間文化研究機構資源共有化システムについて」『シンポジウム　地域研究と情報学

——新たな地平を拓く　講演論文集』107–136.
————. 2008.「空間に基づいた情報解析ツール」『アジア遊学』柴山　守; 原　正一郎; 貴志俊彦（編）, 113: 128–135. 勉誠出版.
原　正一郎; 相田　満; 入口敦志; 江戸英雄; 五島敏芳; 山田直子. 2005.「データベース共有におけるデータマッピングの事例的研究」『研究報告「人文科学とコンピュータ（CH）」』2005-CH-67: 31–38. 情報処理学会.
原　正一郎; 柴山　守. 2007.「地域情報学の構築と時空間情報解析ツール」『人文科学とコンピュータシンポジウム論文集』2007(15): 71–78. 情報処理学会.
原　正一郎; 柴山　守; 安永尚志. 2003.「メタデータによるデータベースの機関間連携の実現——人文科学データ共有のための標準化」『人文科学とコンピュータシンポジウム論文集』2003(21): 17–22. 情報処理学会.
原　正一郎; 安永尚志. 2000.「国文学電子資料館システム」『国文学研究資料館紀要』26: 25–52. 国文学研究資料館.
HL7. 2007. *HL7 Version 3*. http://www.hl7.org/.
International Council on Archives. 1999. *ISAD(G): General International Standard Archival Description Second Edition*. http://www.ica.org/sites/default/files/isad_g_2e.pdf.
The International Committee for Documentation of the International Council of Museums (ICOM-CIDOC). 2009. http://www.cidoc.icom.org.
ISO. 1986. *ISO 8879: 1986 Information Processing—Text and Office Systems—Standard Generalized Markup Language (SGML)*.
————. 1987. *ISO 9745, Electronic Data Interchange for Administration, Commerce and Transport (EDIFACT)*.
————. 1999. *ISO/IEC 13250 Topic Maps Information Technology Document Description and Processing Languages*. http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf.
国立国会図書館. 2009.「デジタルアーカイブポータル（PORTA）」, http://porta.ndl.go.jp/portal/dt.
国立情報学研究所. 2009. http://www.nii.ac.jp
The Library of Congress (LC). 2002. *Encoded Archival Description Version 2002 Official Site*. http://www.loc.gov/ead/index.html
————. 2004a. *SRU (Search/Retrieve via URL)*. http://www.loc.gov/standards/sru/.
————. 2004b. *MARC 21 Format for Bibliographic Data National Level Record - Bibliographic Full Level & Minimal Level*. http://www.loc.gov/marc/bibliographic/nlr/nlr.html#intro.
————. 2009a. *MODS Metadata Object Description Schema Official Web Site*. http://www.loc.gov/standards/mods/.
————. 2009b. *Metadata Encoding & Transmission Standard*. http://www.loc.gov/standards/mets/mets-schemadocs.html.
溝口理一郎. 2005.『オントロジー工学』オーム社.
桶谷猪久夫. 2007.「人文分野における日本地名辞書の構築と地名属性の特徴分析」『人文科学とコンピュータシンポジウム論文集』2007(15): 79–86. 情報処理学会.
OCLC. 2009. http://www.oclc.org.
関野　樹. 2008.「時間に基づいた情報解析ツール」『アジア遊学』柴山　守; 原　正一郎; 貴志俊彦（編）, 113: 140–148. 勉誠出版.
関野　樹; 久保正敏. 2007.「T2Map——時間情報に特化した解析ツール」『人文科学とコンピュータシンポジウム論文集』2007(15): 183–188. 情報処理学会.
柴山　守; 原 正一郎. 2008.「総論　地域情報学の目指すところ——地域情報学における GIS の応用」『アジア遊学』柴山　守; 原　正一郎; 貴志俊彦（編）, 113: 28–35. 勉誠出版.
W3C. 2008. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. http://www.w3.org/TR/xml/.

付録1
人間文化研究機構資源共有化システムの時空間メタデータの定義

（補足）
- 付録1はNIHUシステムの仕様の一部であり，Coverage要素内の拡張した部分のみを示している。概要を示すことが目的であるため，ここではDTDやXML Schemaではなく，視覚的に図示している。
- 実際のNIHUシステムで使用されている時空間メタデータは，実装の都合で，付録1を簡素化したものになっている。

1．名前空間定義
xmlns:ts＝“http://www.nihu.jp/ResourceSharing/GeoTemporal”

2．Coverage内の記述

```
ts:Container＋//時空間記述単位
 ├ts:Date//時間データ
 │ ├ts:From
 │ │ ├ts:DescDate?//記述時間（和暦など）
 │ │ │ ├ts.description*//元データの記載時間であるが，原則として暦日を記載
 │ │ │ │  DataType＝#PCDATA
 │ │ │ ├ts:era*//～時代・～時代前期・～世紀など暦日以外の時間記述
 │ │ │ │  DataType＝#PCDATA
 │ │ │ ├ts:regionalCal?//区分された時間記述（和暦など）
 │ │ │ │  Attribute＝ts:desc_format [JP]//記述形式
 │ │ │ │    //JP 和暦（デフォルト）
 │ │ │ │ ├ts:gengo?//元号
 │ │ │ │ │  DataType＝#PCDATA
 │ │ │ │ │  Attribute＝ts:desc_precise [accurate, about, maybe]
 │ │ │ │ │    //accurate 正確（デフォルト）
 │ │ │ │ │    //maybe…カ
 │ │ │ │ │    //about…頃
 │ │ │ │ ├ts:year?//年
 │ │ │ │ │  DataType＝#PCDATA
 │ │ │ │ │  Attribute＝ts:desc_precise [accurate, about, maybe]
 │ │ │ │ │    //accurate 正確（デフォルト）
 │ │ │ │ │    //maybe…カ
 │ │ │ │ │    //about…頃
 │ │ │ │ ├ts:month?//月
 │ │ │ │ │  DataType＝#PCDATA
 │ │ │ │ │  Attribute＝ts:desc_precise [accurate, about, maybe]
 │ │ │ │ │    //accurate 正確（デフォルト）
 │ │ │ │ │    //maybe…カ
 │ │ │ │ │    //about…頃
 │ │ │ │ └ts:day?//日
 │ │ │ │    DataType＝#PCDATA
 │ │ │ │    Attribute＝ts:desc_precise [accurate, about, maybe]
 │ │ │ │      //accurate 正確（デフォルト）
```

//maybe…カ
//about…頃
└ts:westernCal?// 西洋（太陽）歴は yyyy-mm-ddThh:mm，ユリウス通日は整数型
DataType = #PCDATA
Attribute = ts:desc_format [JC, JD, GC]//記述形式
//JC ユリウス歴
//JD ユリウス通日
//GC グレゴリアン歴
Attribute = ts:norm_format [ISO, JIS, W3C]//正規化形式
Attribute = ts:timeZone CDATA//国際標準子午線からの時差
└ts:NormDate//正規化時間　西洋（太陽）歴は yyyy-mm-ddThh:mm，ユリウス通日は整数型
DataType = #PCDATA
Attribute = ts:norm_format [ISO, JIS, W3C]//正規化形式
//JC ユリウス歴
//JD ユリウス通日
//GC グレゴリアン歴（デフォルト）
Attribute = ts:norm_format [ISO, JIS, W3C]//正規化形式
Attribute = ts:timeZone CDATA//国際標準子午線からの時差
Attribute = ts:desc_precise [accurate, about, maybe]
//accurate 正確（デフォルト）
//maybe…カ
//about…頃
└ts:To
├ts:DescDate?//記述時間（和暦など）
├ts.description*//元データの記載時間であるが，原則として暦日を記載
DataType = #PCDATA
├ts:era*//～時代・～時代前期・～世紀など暦日以外の時間記述
DataType = #PCDATA
├ts:regionalCal?//区分された時間記述（和暦など）
Attribute = ts:desc_format [JP]//記述形式
//JP 和暦（デフォルト）
├ts:gengo?//元号
DataType = #PCDATA
Attribute = ts:desc_precise [accurate, about, maybe]
//accurate 正確（デフォルト）
//maybe…カ
//about…頃
├ts:year?//年
DataType = #PCDATA
Attribute = ts:desc_precise [accurate, about, maybe]
//accurate 正確（デフォルト）
//maybe…カ
//about…頃
├ts:month?//月
DataType = #PCDATA
Attribute = ts:desc_precise [accurate, about, maybe]

```
        //accurate 正確（デフォルト）
        //maybe…カ
        //about…頃
    ts:day?//日
      DataType＝#PCDATA
      Attribute＝ts:desc_precise [accurate, about, maybe]
        //accurate 正確（デフォルト）
        //maybe…カ
        //about…頃
  ts:westernCal?//西洋（太陽）歴は yyyy-mm-ddThh:mm，ユリウス通日は整数型
    DataType＝#PCDATA
    Attribute＝ts:desc_format [JC, JD, GC]//記述形式
        //JC ユリウス歴
        //JD ユリウス通日
        //GC グレゴリアン歴
    Attribute＝ts:timeZone CDATA//国際標準子午線からの時差
    Attribute＝ts:norm_format [ISO, JIS, W3C]//正規化形式
ts:NormDate//正規化時間 西洋（太陽）歴は yyyy-mm-ddThh:mm，ユリウス通日は整数型
  DataType＝#PCDATA
  Attribute＝ts:desc_format [JC, JD, GC]//記述形式
        //JC ユリウス歴
        //JD ユリウス通日
        //GC グレゴリアン歴（デフォルト）
  Attribute＝ts:norm_format [ISO, JIS, W3C]//正規化形式
  Attribute＝ts:timeZone CDATA//国際標準子午線からの時差
  Attribute＝ts:desc_precise [accurate, about, maybe]
        //accurate 正確（デフォルト）
        //maybe…カ
        //about…頃
ts:Place//位置データ
  ts:DescPlace?//記述位置（住所など）
    ts:description＋
    DataType＝#PCDATA//元データの記載地名
    Attribute＝ts:desc_format [LL, AD, ZIP, NM, TEL, CD]//記述形式
        //LL　緯度および経度　この場合“緯度，経度”で記述する
        //AD　住所
        //ZIP　ZIP コード
        //NM　地名
        //TEL　電話番号
        //CD　コード
  ts:address*//区分された地名記述（住所）
    ts:nation?//国家・地域名
      DataType＝#PCDATA
      Attribute＝ts:desc_precise [accurate, about, maybe]
        //accurate 正確（デフォルト）
        //maybe…カ
```

```
        //about…頃
    ├ts:prefecture?//県・旧国名
      DataType＝#PCDATA
      Attribute＝ts:desc_precise [accurate, about, maybe]
        //accurate 正確（デフォルト）
        //maybe…カ
        //about…頃
    ├ts:county?//郡・市
      DataType＝#PCDATA
      Attribute＝ts:desc_precise [accurate, about, maybe]
        //accurate 正確（デフォルト）
        //maybe…カ
        //about…頃
    └ts:town?//町名以下
      DataType＝#PCDATA
      Attribute＝ts:desc_precise [accurate, about, maybe]
        //accurate 正確（デフォルト）
        //maybe…カ
        //about…頃
└ts:NormPlace//正規化位置
  ├ts:description＋//地名，ZIP コード，電話番号，コード
    DataType＝#PCDATA
    Attribute＝ts:desc_format [LL, ZIP, NM, TEL, CD]//記述形式
        //LL　緯度，経度　この場合“緯度，経度”で記述する
        //ZIP　ZIP コード
        //NM　地名
        //TEL　電話番号
        //CD　コード
  ├ts:address?//区分された地名記述（住所）
    ├ts:nation?//国家・地域名
      DataType＝#PCDATA
    ├ts:prefecture?//県・旧国名
      DataType＝#PCDATA
      Attribute＝ts:desc_precise [accurate, about, maybe]
    ├ts:county?//郡・市
      DataType＝#PCDATA
    └ts:town?//町名以下
      DataType＝#PCDATA
  └ts:coordinate
    Attribute＝ts:CoordinateSystem［ArcGIS に準拠］//座標系
    ├ts:NWpoint//矩形北西点
      ├ts:x//緯度あるいはＸ座標
        DataType＝#PCDATA
      └ts:y//経度あるいはＹ座標
        DataType＝#PCDATA
    └ts:SEpoint//矩形南東点
```

```
        ├ts:x//緯度あるいはX座標
        │  DataType＝#PCDATA
        └ts:y//経度あるいはY座標
           DataType＝#PCDATA
├ts:Who?//人物データ
│ ├ts:DescWho?//記述人物名
│ │  DataType＝#PCDATA
│ └ts:NormWho?//正規化人物名
│    DataType＝#PCDATA
└ts:Event?//事象（what，how に相当）
   DataType＝#PCDATA//事象記述
```

注　　釈

注1）メタデータ（metadata）は「データのデータ」という意味であるが，ここではデータベースのレコードを指す。例えば書誌データベースの場合，本というデータを検索するためのレコードがメタデータとなる。注3）の関係データベースの場合，1レコードを構成する各フィールドの名称・出現順序・データ型（例えば文字列・整数）などを定義することが，メタデータの定義となる。記述の粒度（例えば住所はどの程度まで詳細化するか），記述規則（例えば県名と郡市を区別するなら，どのように記述するか），言語（例えば日本語），符号化法（例えばユニコード）などの検討も，メタデータの設計においては重要である。

注2）文書はヘッダ，タイトル，章，節，文，単語，文献，謝辞などの部品から成り立っている。文書構造は，どのような部品がどのように組み合わされて1つの文書を構成しているかを示している。例えば，「①ある文書は1つのヘッダと1つ以上の章から構成され，②ヘッダは1つの文書タイトルと1つ以上の著者名と所属の対およびアブストラクトから構成され，③章は1つの章タイトルと1つ以上の段落から構成され，④段落は1つ以上の文から構成され，⑤文には0個以上の地名，人名，時間に関する単語が含まれている」は，文書構造の例である。

構造化がそれほど高度ではない文書の例として，HTML（Hyper Text Markup Language）文書をあげることができる。文書中の段落程度は指定できるが，人名や地名や時間などの単語を指定することはできない。

注3）これは関係データベース（relational database）に基づいた説明である。関係データベースを簡単に述べると，Microsoft Access や MySQL などのように，データをテーブル（あるいは表）形式で表現しているデータベースである。関係データベースの構造は，

- 表全体をファイル（file）と呼ぶ
- ファイルは1つ以上の行から構成され，1行が1つの情報単位であり，レコード（record）と呼ぶ
- レコードは1つ以上の列から構成され，1列が1つの属性（attribute）であり，フィールド（field）と呼ぶ

のようになっている。もしデータベースが書誌であれば，ファイルは書誌データ全体，1レコードは1冊の本に関する書誌データ，列はその本の題名，主題，著者名，出版社，出版年などの属性に対応する。つまりフィールドはデータの意味（semantics）を表しているとみなすことができる。なお関係データベースのデータ操作や定義を行うための体表的なデータベース言語に SQL がある。

注4）インターネット上のデータを何らかの方法で収集できるものとする。しかし収集されるデータの構造，つまりデータの項目前・型・順序・繰り返し回数・記述言語などは，それぞれ異なっているのが一般的である。例えばA医学図書館の OPAC では書名に関するデータ項目を「title」という名前で識別しているが，B医学図書館書では「shomei」であったとする。ここで2つの OPAC から「診断と先端技術: 開発と応用の現状をさぐる」という本を，SQL という一般的な検索言語を使って検索する場合，A医学図書館の OPAC なら「select * from OPAC_A where title＝“診断と先端技術: 開発

と応用の現状をさぐる”」となる検索式も，B医学図書館なら「select * from OPAC_B where shomei＝“診断と先端技術: 開発と応用の現状をさぐる”」となる。このようにデータベースごとに検索命令が異なると，ネットワークから効率的に情報を探し出すのは困難である。さらにC医学図書館では書名が「title」と「subtitle」という２つのデータ項目に分けられ，title に「診断と先端技術」，subtitleに「開発と応用の現状をさぐる」が登録されているものとする。こうなると，上記と同様の検索命令「select * from OPAC_C where title＝“診断と先端技術: 開発と応用の現状をさぐる”」では，本を見つけることさえできなくなってしまう。

注５）Web には表層 Web と深層 Web の２種類があると言われている。表層 Web は一度作成されると変更されることの少ない HTML で書かれた静的ページなどで，いわゆるホームページに相当する。深層 Web の実態はホームページからリンクされているデータベースであり，検索されるたびに動的に生成されるページである。

表層 Web の情報収集には，検索エンジンと呼ばれるソフトウェアを用いる。これにもロボット型探索エンジンと登録型検索エンジンの２種類がある。ロボット型探索エンジンは Web ページを１ページずつ調べ，そこに設定されている全てのハイパーテキストリンクを辿りながら，次々に探査するページの範囲を広げていくのものである。この典型例が Google である。登録型検索エンジンではタイトルなどをあらかじめ登録しておき，その登録されたホームページの中から検索を行うものである。手作業であるためロボット型検索エンジンよりも登録されるページ数は少ないが，高い検索精度を期待できる。この典型例が Yahoo である（ただし現在の Yahoo は登録型とロボット型両方の機能を持ち合わせている）。検索エンジンと異なった方法に，サブジェクトゲートウェイ（subject gateway）がある。これはネットワーク上のデータを定められた基準に従って収集し，（人手による精度管理を行った上で）構造化されたメタデータとして整理して検索できるようにしたものである。これにより，価値の高い情報に効果的にアクセスすることが可能となる。

深層 Web はデータベースあり，データ構造や検索法がデータベースごとに異なるため，通常の検索エンジンでは利用できない（注４）。このような場合，データ提供側のコンピュータ（サーバ）と受容側のコンピュータ（クライアント）間で，交換される検索命令・データの種類・形式を前もって決めておき，これに従ってデータ交換を行う必要がある。コンピュータ間のデータ交換に関する様々な取り決めを規約（プロトコル: protocol）と呼ぶ。資源共有化システムは，深層 Web を対象とした情報検索システムの１つと見ることができる。

注６）テキストの一部をタグ（tag）と呼ばれる特別な文字列で囲むことにより文書の構造（注２）を明示する方法をマークアップ（markup），マークアップのための言語をマークアップ言語（markup language）という。よく知られたマークアップ言語にはホームページ用の HTML（Hyper Text Markup Language）がある。また汎用的な言語としては SGML（Standard Generalized Markup Language），さらに SGML から発展した XML（eXtensible Markup Language）などがある。例えば本稿のⅡ章は，XML を利用して，

```
〈Chapter No=“2”〉〈Title〉資源共有化システムの概要〈/Title〉
  〈Section〉資源共有化システムは，地域研究に…〈/Section〉
  〈SubChapter〉〈Title〉資源共有化システムの枠組み〈/Title〉
    〈Section〉地域研のデータベースは，…〈/Section〉
      …
  〈/SubChapter〉
  〈SubChapter〉〈Title〉資源共有化システム開発の経緯〈/Title〉
    〈Section〉資源共有化システムの開発は…〈/Section〉
      …
  〈/SubChapter〉
  …
〈/Chapter〉
```

のように書くことができる。

〈Title〉などのように「〈」と「〉」で囲まれた文字列をタグと呼ぶ。タグには，〈Title〉のように「〈」に通常の文字列が続く形式の開始タグと，〈/Title〉のように「〈/」に開始タグと同じ文字列が続く形式の終了タグがある。開始タグと終了タグに囲まれた部分が，そのタグで示された文書構造の領域であり，要素あるいはエレメント（element）と呼ぶ。また開始タグと終了タグの間の文字列が，その文書構造のデータ内容あるいは値となる。例えば文書中のある文字列部分を開始タグ〈Title〉と終了タグ〈/Title〉で囲み，「〈Title〉資源共有化システム開発の背景と経緯〈/Title〉」となっている場合，その部分が Title 要素（章のタイトル）であり，その内容あるいは値が「資源共有化システム開発の背景と経緯」であることを示している。

注7）Z39.50 はネットワーク環境下における検索質問・検索結果・認証など，情報検索システムに必要な機能を定義した国際標準規約である。1970年代に米国議会図書館と書誌ユーティリティとの間で，コンピュータに蓄積されていた書誌データを直接交換しようとする計画に端を発している。Z39.50 はデータベースシステムのハードウェアやソフトウェアの実装に依存しない手順を実現している。

Z39.50 はサーバ・クライアント方式の規約であり，サーバ側のデータベースシステムとクライアント側の検索ソフトが Z39.50 の規約に従って情報交換を行う限り，利用者は使い慣れた検索環境下で複数のデータベースにアクセスできる。しかし，Z39.50 が開発された当時の書誌検索システムは大型計算機による集中管理方式が取られていた。つまり Z39.50 はインターネット環境には必ずしも適合しない規約である。

注8）SRU/SRW はデータの検索と取得を想定したインターネット環境に適合した規約である。米国議会図書館の Z39.50 管理維持機関のプログラムの1つとして位置付けられている ZING（Z39.50 International Next Generation）が開発した。データ記述形式は XML でなければならないが，その結果としてデータ加工や処理は，それ以前に比べると容易になり信頼性も高くなる。

SRU と SRW の違いは，SRU が検索要求に REST（Representational State Transfer）フレームワークに基づいた URL を用いているのに対して，SRW は検索要求に SOAP 仕様に基づいた XML を用いる点である。URL を打ち込めば使える SRU は，特殊なクライアントがなくても Web ブラウザからも簡単に利用でき，SRW よりも多くのサービスが利用されている。

注9）定義上，データベースのデータ構造は全てメタデータとなる。しかしデータベースは共有財産であるから，メタデータには「広く利用されるべきである」という意味も含まれている。注4）で述べたように，同じ書誌データベースであってもメタデータが異なると共有化は困難となる。したがって書誌データベースを構築するのであれば，全てのデータベースは普及している書誌用メタデータに従うべきである。このように広く流通しているメタデータを本研究では標準メタデータと呼ぶ。

注10）XML では固有のタグ集合を自由に定義できるが，標準的なタグ集合を利用した方が効率が良く相互運用性も高まる。つまり，1つの XML 文書を作成する際に，既存の複数のタグ集合を組み合わせて利用できるような仕組みがあれば便利である。ところが title のようなタグは書物や肩書など様々な場面で登場する可能性がある。したがって同じ title であっても，どのタグ集合に属しているのかを区別できなければならない。この問題を解決する方法として XML 名前空間（Name Space）がある。

XML 名前空間は，タグ集合を URI と組み合わせる（修飾する）ことで重複するタグ名を識別する。URI は世界で重複することのない識別子であるから，タグセットごとに異なる URI を割り当てるという方法は合理的である。例えば，タグ集合Aを http://example.com/sets/a/ という URI で修飾し，タグ集合Bを http://example.com/sets/b/ という URI で修飾する。両方のタグ集合に「title」がある場合，形式的には {http://example.com/sets/a/}title および {http://example.com/sets/b/}title のように記述すれば両者を区別できる。

しかし全てのタグに URI を記述するのは煩雑である。そこで XML 名前空間では，各 URI に仮の名前を割り当て，それを接頭辞として用いる。例えば http://example.com/sets/a/ に対して「aset」，http://example.com/sets/b/ に対して「bset」という名前を仮に割り当てると，{http://example.

com/sets/a/{title は aset:title，{http://example.com/sets/b/{title は bset:title と書くことができる。

注11）XML の属性（attribute）は，XML 文書の中で要素に対して付加的な情報を付け加えるために使用される。付加される情報は属性名と値の対で表現され，開始タグの中に記述される。例えば

〈img source＝“http://example.com/sample.jpg”〉サンプル画像〈/img〉

の場合，img 要素の開始タグの中に source という名前の属性があり，その値が「http://example.com/sample.jpg」であることを示している。

注12）あるプラットフォーム（OS など）が提供する機能を，外部のソフトウェアが利用する際の手続きに関する約束。OS やミドルウェアは，多くのソフトウェアが共通して必要とする機能を提供している。ソフトウェアの開発者は，そのような機能を自分でプログラミングする必要はなく，ただ約束に従ってその機能を呼び出すだけで，その機能を利用したソフトウェアを作成することができる。

注13）MARC には，仕様が複雑，固定長のタグ付けが面倒，可変長フィールドのタグが数字で表現されていて覚えるのが面倒など，多くの問題が指摘されており，より簡易な記述要素集合の開発が望まれていた。米国議会図書館では MARC のデータ要素と意味構成を基本として 2 つの XML ベースのスキーマを開発した。1 つは MARC の全データ要素を XML 構文に変換した MARC XML である。もう 1 つは MARC のサブセットで構成され，MARC よりも簡単で，数字ではなく言語タグで表現できる MODS（Metadata Object Description Schema）である。MODS は，DCMES のような非常に簡易な形式から MARC21 のような複雑な構造を持つ詳細な形式にまで対応できる特徴がある。

注14）EAD（Encoded Archival Description: 符号化記録史料記述）は，アーカイブの検索手段を電子的に符号化するためのデファクトスタンダードである。図書館における MARC に相当する。

本稿では，アーカイブを広義に定義して図書館や博物館を含むものとする。アーカイブを強調する理由は，資料を群あるいはコレクションとして扱い，それらを生成した個人・組織の活動および周辺情報とともに体系的に整理したいためである。

アーカイブにおいて資料は階層的に整理される。例えば，「ある資料は○○株式会社の△△部の□□課の～～係が作成した」，あるいは「この写真は某先生資料群の第○番目の箱の第△番目の写真アルバムの□ページの～番目」といった具合である。EAD はこの階層構造を記述することができ，Collection，Series，Sun-Series の順に階層化される。

注15）デジタル技術は日進月歩であり，現時点では最新のフォーマットや手法も早晩陳腐化する。デジタルコンテンツを将来も利用可能とするためには，保存対象のビット列を再生する際に必要な技術情報（フォーマット，再生ハードウェア・アプリケーションなどの技術要素）も記述する必要がある。

注16）トピックマップ（Topic Maps）は，ある主題（topic）と主題間の関連（association）および主題と関連する情報資源（occurence）の 3 つの情報を利用して知識を表現する枠組みである。